Contact or inheritance? New evidence on the Proto-Philippines debate

Victoria Chen[1], Kristina Gallego[2], Jonathan Kuo[3], Isaac Stead[4], and Benjamin van der Voorn[5]

(authors listed in alphabetical order)

[1]Victoria University of Wellington, [2]University of the Philippines Diliman, [3]National Taipei University of Technology, [4]Max Plank Institute for Evolutionary Anthropology, [5]Payap University

An outstanding question in Austronesian higher-order subgrouping concerns the linguistic position of Philippine languages. Due to a lack of attention to comparative evidence beyond lexical innovations, it remains unsettled whether these languages diversified from a shared single ancestor excluding all Malayo-Polynesian (MP) languages outside the Philippines or constitute multiple primary branches of MP. We present three lines of new evidence countering previous arguments for Proto-Philippines (PPh) (Blust 2019 et seq.; Zorc 1986, 2020). First, we highlight the absence of PMP *d/z merger in Central Luzon languages, which undermines the sole phonological innovation defining PPh. Second, we examine the semantic categories of PPh-defining lexical items and their geographical distribution, demonstrating that both suggest a high likelihood of borrowing rather than inheritance. Finally, we explore an understudied variation in Circumstantial Voice morphology in Philippine languages, showing new evidence for multiple layers of borrowings across Philippine subgroups. We conclude that the high number of lexical innovations previously proposed as evidence for PPh is better viewed as the outcome of various types of contact scenarios (diffusion, borrowing, and linkage histories), as suggested in Ross (2020), rather than a case of lexical retention.

Keywords: ○ Proto-Philippines ○ Austronesian higher-order subgrouping ○ lexical innovation ○ morphological borrowing ○ language contact ○ linkage

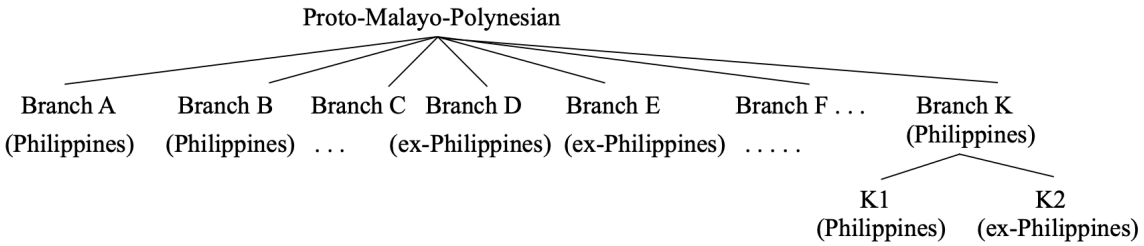
Contents

1 Introduction

Recent debates on the linguistic position of Philippine languages (Blust 2019 et seq.; Zorc 2019; Reid 2020; Ross 2020; Liao 2020; a.o.) have highlighted the limitations of the Neogrammarian approach to subgrouping for classifying closely related languages under extensive contact. Although there is strong comparative evidence that all Malayo-Polynesian (MP) languages spoken outside the Philippines constitute distinct primary branches (Blust 2001; Ross 2004; Smith 2017), the linguistic position of the languages spoken by the in-situ population in the Malayo-Polynesian homeland remains a point of contention. On one hand, very little little phonological evidence suggests their affinity; on the other, more than 1,000 shared lexical items have been identified across these languages (although their distribution varies from one item to another), none of which is attested in MP languages outside the Philippines. Several researchers have thus argued for the existence of a shared common ancestor of all modern Philippine languages – Proto-Philippines (PPh) – as a distinct primary branch of MP (Blust 2019, 2020, 2021; Zorc 2020). This affinity, as proposed in Blust's series of work, is the outcome of a historical levelling event that eliminated other PMP descendants in the Philippines. However, the lack of non-lexical evidence for this subgroup has led to an alternative view that the shared vocabulary is the outcome of lexical diffusion (Reid 2020; Ross 2020; Liao 2020).

From a theoretical point of view, the puzzle arose from the challenge in identifying distinct innovations for the pattern of dispersal of the in-situ population that diversified in the first landing site of a series of inter-island diaspora. As Luzon is adjacent to the Visayas and Mindanao island groups, one would expect a subgrouping scenario like (1), in which PMP speakers gradually diversified into distinct linguistic communities through the Austronesian's expansion across the Greater Philippine region (alongside the migrations out of the Philippine islands). Philippine languages would therefore represent several independent primary branches of PMP, parallel to ex-Philippine branches. It is also possible that some of the ex-Philippines branches derived from one of the branches that diversified in the Philippines island, as exemplified by Branch K.

(1) Hypothetical subgrouping tree of MP higher-order languages



However, the apparent similarities among Philippine languages have led to the superficial yet well-adopted assumption that they all descended from a single protolanguage (Blake 1906; Scheerer 1918; Charles 1974; Llamzon 1975; Paz 1981; Blust 2005, 2019 et seq.; a.o.). This paper contributes to this debate by re-examining current arguments in the recent literature drawing on new data. The data concerns specific aspects of the phonology, lexicon, and morphosyntax of Philippine languages, pertaining to three central questions (2a–c).

(2) a. How robust is the merger of PMP *d and *z attributed to PPh?

b. What history can be reconstructed for the distribution of the proposed lexical innovations across the Philippine subgroups?

c. Can the domain of morphosyntax tell us something new about the relationship of the Philippine languages?

We begin with a literature review of the ongoing debate (section 2). Section 3 explores the reflexes of forms with PMP *d and *z among Philippine languages, particularly the languages of Central Luzon. Section 4 maps selected lexical innovations across Philippine subgroups to identify possible borrowing scenarios, and to examine the distribution of the innovations in terms of semantic fields following Haspelmath and Tadmor's (2009) quantitative study on borrowing. Section 5 examines the derivations of Circumstancial Voice across Philippine languages. Section 6 concludes with the argument that the linguistic features observable among Philippine languages today reflect several layers of change, which involve retentions from PAN/PMP at the deepest level, shared innovations out of a linkage history, and finally, borrowed features from later contact among groups.

2 Proto-Philippines: puzzles and debates

Austronesian languages of the Philippines have long been considered an individual subgroup of Malayo-Polynesian (Blake 1906; Scheerer 1918; Charles 1974; Llamzon 1975; Paz 1981; a.o.). Reid (1978, 1982) was the first to point out the lack of exclusively shared phonological and morphosyntactic innovations for PPh, who then concludes that Philippine languages may not have derived from a common ancestor. In response to that view, Zorc (1986) and Bust (2005) present 327 lexical innovations in support for PPh, with the claim that these replacement innovations constitue strong evidence for a shared common ancestry post-PMP.

Over the past four decades, the major objections against this view have been that the phonological and grammatical systems of PPh do not exhibit such degree of innovation, and still in fact, remain identical to Proto-Malayo-Polynesian. A recent reevaluation of Malayo-Polynesian subgrouping (Smith 2017:472) also argues against the validity of PPh because of the low quality of the posited lexical innovations. Blust (2019) revisits the debate with a substantial addition to the earlier list of innovations, presenting a total of 1,259 lexical innovations and one phonological innovation involving the merger of PMP *z and *d. In a series of commentaries, Zorc (2020) concurs with Blust's claims, whereas Liao (2020), Reid (2020), and Ross (2020) maintain their position against PPh, arguing that the pattern of supposed innovations for PPh involves overlapping distribution across the Philippine subgroups, which is indicative of a linkage history.

A linkage history for the Philippine languages agrees with the claim that there has been a rapid expansion of PMP speakers in island Southeast Asia, which left behind an early dialect network in the Philippines. The rapid expansion of speakers means that there has been no sufficient time for a unitary subgroup of languages to develop (or an innovation-defined subgroup), and instead, it is most likely for languages to form a linkage (or an innovation-linked subgroup), united by an overlapping pattern of innovations (François 2014:170–171; Ross 1995:45–46). Blust (2019:183–184, 2020:453–454) acknowledges that such patterning is indeed observable among the Philippine languages, implying a scenario where the languages do not form a single homogeneous subgroup. However, the main point of contention among the scholars is at what level the Philippine languages are commonly descended. Ross (2020:369) notes that PMP is the protolanguage for this early Philippine dialect network, whereas Blust (2020:452) argues for the existence of PPh, maintaining that the innovations are only found in Philippine languages and not elsewhere in Malayo-Polynesian, especially among languages spoken south of the Philippines.

Current studies making use of evolutionary methods in biology such as Bayesian phylogenetics have also contributed to this debate. Gray et al. (2008) show that while there seems to be no strong evidence for PPh, most of the Philippine languages are grouped together. This likewise points to a linkage history for the Philippines.The two competing hypotheses on the relationship of the Philippine languages essentially differ in the interpretation of the overlapping pattern of innovations. Blust (2020)

argues for a single proto-language that is not dialectally homogeneous to explain the pattern in the data. In contrast, advocates for a linkage history, such as Ross (2020), attribute the innovations to a higher-order proto-language, that is, Proto-Malayo-Polynesian, in which the Philippine languages formed an early dialect chain that soon differentiated after speakers of other Malayo-Polynesian languages left the region.

The debate on the validity of PPh remains unresolved because of how the reconstructed history of a linkage is less certain given that there is a tendency for the innovations to have low evidentiary value compared to those observed in innovation-defined subgroups (Ross, 2020:370). Adding further complexity to the matter is how linguistic signals attributed to linkages tend to be similar to those attributed to language contact, where linguistic features have developed not out of common descent, but from borrowing. It is thus necessary to carefully disentangle the evidence at hand in order to reconstruct a more detailed history for the Philippine languages.

In what follows, we approach the debate through three new lines of evidence. In the next section, we highlight the previously overlooked absence of PMP *d/z merger in Central Luzon languages, which undermines the sole phonological innovation defining PPh. We then examine the semantic categories of PPh-defining lexical items and their geographical distribution, demonstrating that both suggest a high likelihood of borrowing rather than inheritance. Finally, we explore an understudied variation in Circumstantial Voice morphology in Philippine languages, showing new evidence for multiple layers of borrowings across Philippine subgroups.

3 PMP *d/z merger as a PPh innovation? A reappraisal

3.1 Review of Blust 2019

In his 2019 paper, 'The Resurrection of Proto-Philippines', Blust states that the phonological evidence for PPh, though "not robust [...] cannot be ignored" (Blust 2019: 156). This evidence consists of a proposed merger between *z and *d. Blust asserts that no language in the PPh subgroup maintains the distinction between *z and *d in stable forms (e.g. PAN *zalan 'path, road, and PAN *duSa 'two). Consequently, this merger of *z and *d helps to define PPh as a unified subgroup.

He cites one exception to the merger: In Ayta Abellen, a language of the CL family, there are present both *udan* 'rain' (< PAN *quzan) and *ula* 'shrimp' (< PAN *quda). However, he notes that this instance is ambiguous, because "other words [in Ayta Abellen] show a merger of *z and *d word initially, as PAN *depa > *depah* 'armsbreadth' and PAN *zaRami > *dayami* 'rice husk"' (Blust 2019: 156–157).

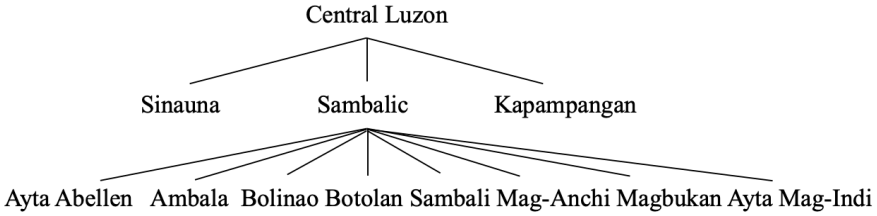
However, further investigation into the ACD reveals many more such exceptions, not only in Ayta Abellen, but also in other Central Luzon (CL) languages. These exceptions are not sporadic or random they reveal a pattern. In the CL languages presented, the reflexes of *z are either /d/ or /r/, depending on the conditioning environment. These reflexes can be explained by regular and uncomplicated sound changes. In the case of *d, however, the reflexes are split. In some instances, a reflex will appear in a form identical to the regular reflex of *z, whereas in other cases a reflex will appear with in the divergent form /l/. If *z and *d were already merged in Proto-CL, then we would expect the patterning of the reflexes to be identical for *z and *d, but that is not the case. This difference in reflex patterning strongly suggests an original separation between *z and *d in Proto-CL[1], undermining the viability of the *d/z merger as a defining sound change for PPh.

[1]This is not a new observation. Although he does not make a connection to the PPh debate, Himes (2012, pg. 492) independently concludes that *z and *d were originally unmerged in Proto-CL. See section 3.4.1 for more details.

3.2 Overview of Data

There are three lower-order branches within CL, which are Kapampangan, Sambalic, and Hatang Kayi[2]. Out of those subgroups, there were four languages for which a useful amount of data was readily available. These are Kapampangan (KP), Ayta Abellen (AA), Ayta Mag-Antsi (AM) and Botolan (BT).[3] The interrelationship of these languages is illustrated in (3). The latter three languages are part of the Sambalic subgroup. Due to a lack of ready data, the Hatang Kayi branch is not represented in this paper.

(3) Central Luzon subgrouping



We collected and reviewed all the reflexes of *z and *d available in the ACD for Kapampangan, Ayta Abellen, Ayta Mag-Antsi, and Botolan. The data was sourced from Blust and Trussels ACD, which can know be found at https://acd.clld.org/. All cited data is reproduced from the ACD entries, so some words will be listed as descending from PPh, despite the fact that this paper is arguing against its existence. The figure below displays a summary of the data: in the second column, the regular sound change rules which affect both *z and *d in each language; in the third column, the instances where *d diverges and is realized as /l/. The key generalizations of the figure are outlined in (4).

(4) a. In all four languages, *z is realized as /d/ word-initially. Word medially, *z is realized as /d/ in Ayta Abellen, while in the other three languages it is realized as /r/.

b. In many instances, *d is realized identically to *z, as either /d/ or /r/.

c. But contra Blust, instances of exceptional *d > l are found in many stable forms such as PAN *daNum > AA *lanom* 'water', PAN *duSa > BT *lua* 'two', and PMP *dateng > AA *lateng* 'to arrive; come'.

d. The *d > l sound change occurs vowel-medially in Kampangan. In the Sambalic languages, it occurs more freely: vowel-medially and word-finally in Ayta Mag-Antsi, and in all environments in Ayta Abellen and Botolan. In section 3.4.1, we present a hypothesis that the sound change *d > l / V_V occurred in Kapampangan and the Sambalic languages, while *d and *z were unmerged, and the flow-on effects from this sound change resulted in the split reflexes of *d.

e. In the Ayta Mag-Antsi data, *z and *d appear wholly unmerged. However, there are no instances in that data of word-initial *z. If Ayta Mag-Antsi is in line with the other CL languages examined in this paper, *z and *d will show partial merging word-initially.

[2] Also known by the exonyms Sinauna and Remontado.

[3] More data would be valuable, particularly more instances of *z. A full dictionary of Ayta Mag-Antsi is already available at https://philippines.sil.org/resources/online$_r$esources/sgb, produced by SIL. However, its contents have not yet been sorted and indexed to their proto-forms.

(5)

Language	Identical reflexes of *d/z	Instances of *d realized as /l/
Kapampangan	*d > r / V_V (4 instances) *d > d (46 instances) *z > r / V_V (5 instances) *z > d (5 instances) (64 instances total)	PMP *badas > *balás* 'sand' PMP *ida > *ila* '3PL' PWMP *ludem > *ma-lúlam* 'cloudy, about to rain' PAN *tuduq > *tulu(ʔ)* 'to drip, flow, spill'
Ayta Abellen	*d > d (30 instances) *z > d (5 instances) (47 instances total)	PAN *daNum > *lanom* 'water' PMP *dateng > *lateng* 'to come' PWMP *sidem (or *silem) > *hilem* 'afternoon' PAN *dengeR > *lenge* 'to hear' PMP *dingding > *lingling* 'wall of a house' PWMP *di hipaR > *lipay* 'the other side of a body of water' PAN *duSa > *lowa* 'two' PAN *qañud > *anol* 'to be carried on the current' PPh *sápad > *sapal* 'hand of bananas' PPh *tadék > *talek* 'dance; to dance' PAN *tuduR > *toloy* 'to sleep' PWMP *pu(n)dut (or *pu(n)zut) > *polot* 'to pick up'
Ayta Mag-Antsi	*d > r / V_V (1 instance) *d > d (1 instance) z > r / V_V (3 instances) (9 instances total)	PMP *badas > *balah* 'sand' PMP *qalad > *alal* 'fence' PAN *qudang > *ulang* 'squid' PWMP *tadu > *talo* 'beeswax'
Botolan	*d > r / V_V (4 instances) *d > d (17 instances) *z > r / V_V (2 instances) (33 instances total)	PPh *agud > *ágol* 'to moan, agonize' PAN *daNum > *lánum* 'water' PAN *duSa > *lua* 'two' PAN *SateD > *atel* 'escort' PPh *tadék > *talék* 'dance; to dance' PWMP *tadu > *tálo* 'beeswax' PMP *tidaq > *tíla*ʔ'remainder' PMP *tudung > *tolóng* 'head cover' PAN *tuduR > *túluy* 'to sleep' PAN *tuduS > *tóʔol* 'knee'

3.3 Proto-Philippines bereft of phonological evidence

At first glance, the realizations of *d and *z in the sampled CL languages seem mixed and unpredictable, but a closer look reveals that *d and *z are realized according to their own distinct patterns. First, *z is always realized *consistently*[4], appearing as /d/ in one environment and /r/ in another. Second, *d is realized *inconsistently* – in an identical environment, it can either merge with *z, appearing as /d/ or /r/ respectively, or it can appear as /l/. Now, if *d and *z were merged at the level of Proto-CL, we would expect to see the same pattern of realization in their reflexes throughout all the Proto-CL descendant languages. However, that is not what we see. In all four CL languages for which sufficient data is available, the reflexes of *d show an inconsistent pattern of realization, while the reflexes of *z show a consistent pattern. The conclusion is that *d and *z are in fact unmerged in these languages, strongly suggesting an underlying separation of *z and *d to the level of Proto-CL. With this new data in view, there is no longer any firm foundation to establish the *d/z merger as a defining sound change of Proto-Philippines, leaving it with no phonological support.

Two objections will occur against the data presented: first, that the inconsistency in the realization of *d is a result of loans from other Philippine languages; second, that the inconsistency is a result of

[4]PWMP *puzut > AA *polot* seems to be an exception. But *puzut has a disjunct form, *pudut, both meaning 'to pick up with the fingers'. If *polot* is a reflex of *pudut there is no inconsistency.

sporadic change. The first of these objections is the stronger of the two, as the CL languages, particularly Kapampangan, have been influenced by surrounding languages – for example Tagalog. Due to the overall phonological similarity of Philippine languages, it can also be difficult to identify loanwords from one Philippine language to another. However, both objections fail in the face of the data on *z. If *z and *d were merged in Kapampangan, and the inconsistency in the reflexes of *d is solely a result of borrowing and/or sporadic change, statistically speaking we ought to see a comparable level of inconsistency in the realization of *z. On other words, if *z and *d had already merged into a united phoneme in Proto-CL, then we would expect words with *z or *d to be replaced by borrowing and/or sporadic change at roughly equal rates. But this is not what we see - all the reflexes of *z across the four languages studied are realized according to a regular rule, while 23% of the reflexes of *d (71% if only counting vowel-medial reflexes) appear as the divergent reflex /l/[5]. These facts strongly suggest that the difference in the realizations of *z and *d is not the result of borrowing or sporadic change, but the result of an underlying division between *z and *d in Proto-CL.

3.4 Additional notes

3.4.1 A hypothesis on the source of split PMP *d

Himes (2012) attempts to create an account of the CL languages using regular sound change rules. One of his proposed rules is *d > l / V_V in the Sambalic languages and Kapampangan. As well as this, he proposes a number of other rules to explain the split realization of *d, but these are lacking in that they are highly targeted rules which often lack phonetic motivation, and that they leave much of the data still unexplained.[6] We suggest that the sound change *d > l / V_V is sufficient to explain not only the vowel-medial instances of *d > l, but also the word-initial and word-final instances, once the possibility of back-formation from affixed forms or forms in connected speech is taken into account.

First we examine the straightforward instances of *d > l / V_V. There were seventeen instances of this sound change found in the data, across all four languages, and seven exceptions. The exceptions are PWMP *ledek to pound grain w/ mortar and pestle > AA *ledek*; PPh *pidek > BT/AM *kiṟip* 'eyelash'; PAN *kuden > KP *kúran* 'large cooking pot for rice', PPh *dúdun > KP *durún* 'locust', and PPh *katúday > KP *kature* 'a plant: Sesbania grandiflora'; and PMP *pudul > BT *poról* 'blunt, dull', plus one exception found in Himes (2012) but not the ACD data: PMP *ludaq > BT *ludáʔ* 'spit'. For six of these exceptions, we have good reason to suspect that *d > l / V_V sound change did not apply to them. *ledek* is likely a new coinage from the root *-dek$_2$. *kiṟip*, *durún* and *kature* are all words unique to the Philippines and may be considered as loanwords or new coinages. In the case of *ludá* and *poról*, its worth noting that in their proto-forms, each *d is either followed or preceded by a liquid. It is possible that proximity to a liquid makes the change of *d > l / V_V uncertain, perhaps due to the difficulty of pronouncing one liquid after another, or pressure towards differentiation of phonemes. In other instances, the change goes ahead in such a context: see PWMP *ludem > KP *ma-lúlam* 'cloudy, about to rain' and PAN *tuduR > BT *túluy* 'to sleep'. If the given explanations hold, only the Kapampangan reflex *kúran* from PAN *kuden remains as an unexplained exception to Himes proposed rule, leaving it standing on solid ground.

In the Sambalic languages (but not Kapampangan), *d can also appear as /l/ in word-initial or -final position, though it shows up less frequently than in the vowel-medial position. E.g. In Ayta Abellen, PWMP *damay > AA *damay* 'sympathy, help' but PAN *daNum > AA *lanom* 'water'. We hypothesize that such a split realization of *d could be the result of a secondary effect of the rule *d > l / V_V, that

[5]There are 20 total instances of *z in the data. There are 131 instances of *d, 30 of which appear as /l/.

[6]For example, Himes suggests that in the Sambalic languages, *d > /l/ word initially, except when the next consonant is a liquid. However, there are many stable forms, stretching back to PWMP and earlier, which violate this rule. E.g. PWMP *damay > AA *damay* 'sympathy, help', PAN *depah > AA *depah* 'armsbreadth', or PAN *daqaN > BT *daʔan* 'old (of objects)'. Additionally, there is no clear phonetic motivation for *d to change to /l/ in word-initial but not word-final position.

is, the result of back-formation from affixed forms, or even from forms in connected speech. In the case where *d is in word-initial or -final position, the presence of an affix with no coda would place *d in a vowel-medial position, making it subject to the rule *d > l / V_V. This would result in two competing forms of the same stem being present in the language, one (an l-stem) contained within the affixed form, the other (a d-stem) within the unaffixed stem. For example, if PCL *dateng 'to come' were affixed to become *ka-dateng-an, the affixed form would change to *kalatengan under the rule *d > l / V_V, and later that form could be reinterpreted as consisting of a stem *lateng affixed with *ka- -an. If this process occurred in Proto-Sambalic, it would mean that in some cases, the innovative l-stem would replace the original, and in other cases, the original d-stem would remain, and be re-affixed to push out the innovative affixed form. A similar scenario could play out if a word without a final consonant precedes a word beginning with *d in connected speech, or a word with a final *d is followed by a word without an initial consonant. Phonetically, *d would be an intervocalic position, and thus be realized as /l/. Again, the result would be two competing forms in the language, one with a d-stem and one with an l-stem. Such a scenario would lead to the inconsistent realisation of *d in word-initial and -final position that we see in the Sambalic languages. After this process, the remaining d-stems reflexes would merge with *z, splitting with the l-stems in their realization.

3.4.2 Negrito languages

It is worth noting that a handful of languages discussed above – Ayta Abellen, Ayta Mag-Antsi, Magindi, Magbukun, and Sinauna – are Negrito languages. The archaeological evidence suggests that contact between the Negritos and Austronesian speakers happened early on (Reid 1994), implying that the maintained contrast of *z and *d represented in these languages originates from an early stage of the settlement of the Philippines (closer to PMP), as suggested in Reid (1986). Negrito languages are also on the cultural and physical periphery in the Philippines, and are less likely to be subject to diffusionary forces through the dialect chain. If the Central Luzon languages have remained in situ since the original Austronesian settlement, as seems likely, it is possible they have retained features which diffusion has obliterated from other Philippine languages. On the other hand, no other Negrito language that we have investigated, or any other Philippine language, has so far shown any split in *d and *z.

3.4.3 A path forward for Proto-Philippines?

If the data and conclusions given here are accepted, there may still be a way to maintain, or even increase, the phonological evidence for Proto-Philippines. The CL family is looking increasingly phonologically unique, as it is the only subgroup in the proposed PPh family which preserves both the *d/z merger and the *n/ñ merger – the latter in Kapampangan (see PMP *buñi > *buñi* 'celebrated, acclaimed', and others), and perhaps also Sinauna (see PMP *ñamuk > *yamuk* 'mosquito'). If it could be demonstrated that Central Luzon does not group with the rest of the Philippine languages – for example, if it represents a sister branch of PPh from PMP – then both the *d/z and *n/ñ merger would stand as phonological evidence for PPh. Such a conclusion would go against the consensus on the subgrouping of PPh up until now, and would require demonstrating that it is sufficiently divergent from the rest of the proposed PPh languages – not only phonologically, but also lexically and grammatically. If this is unable to be demonstrated, the persistence of the *d/z distinction in the CL languages remains firm evidence against a unitary PPh.

4 Revisiting lexical innovations that define PPh

We turn now to a reconsideration the lexical evidence which has been proposed for PPh. A key part of this evidence is the list of 1,259 lexical items which have no external cognates outside the Philippines (Blust 2020). It is claimed that a) these lexemes are present in such quantity that they imply descent from a single protolanguage, and b) that a core of them (37 items) represent the strongest type of lexical evidence for common descent, i.e. replacement innovations. Therefore, they must be descended from a Proto-Philippines, which spread throughout the archipelago replacing the descendant languages of Proto-Malayo-Polynesian which had been spoken there since the initial MP expansion. We argue that the distribution of these etyma does not constitute sufficient evidence for common descent, as it is not possible to distinguish borrowings from etyma inherited from a common ancestor. In addition, it is highly likely that a dialect linkage formed during the initial MP expansion, and even widely distributed "PPh" etyma may in fact be items which spread through this dialect network early on.

Much robust discussion has already taken place on the lexical evidence for PPh (see, for example, Smith 2017; Reid 2020; Ross 2020; Zorc 2020). In what follows, we summarise the key arguments against the lexical data as evidence for PPh and expand on these with some new findings. A key issue is that of distinguishing loanwords from common inheritance, and some mention of this is made in each of the papers cited above.

Several authors (Smith 2017; Ross 2020) have already mentioned short and long-distance trading links as a vector for the diffusion of loanwords through the archipelago. Smith (2017: 464), in his review of the strength of Blust's proposed PPh etyma, points to the word for Manila hemp (*Musa textilis*), an important resource for the production of fibre for rope and weaving, as being widely distributed (in the Batanes, Northern Luzon, Central Luzon and Greater Central Philippine microgroups) and as fitting Blust's criteria for a strong witness to PPh. However, Smith suggests that since Manila hemp is a widely traded item, this may be a loan. Blust (2019: 215) counterargues that since Manila hemp is native to the Philippines, PPh speakers presumably had a word for this item, and as such it cannot be a loanword. This issue is also highlighted by Liao (2019).

This example is representative of the debate over the lexical evidence for PPh as a whole. We would make two important points here: first, assuming that speakers of some ancestral Philippine protolanguage had a term for Manila hemp which was retained by its daughter languages, it is in principle not possible to distinguish whether this language was PPh or whether these lexemes arose in and spread through the Proto-Malayo-Polynesian dialect network. This is because, as we have shown, PPh is no longer supported by any regular sound changes. In both the retention or diffusion scenarios the simplest explanation excludes the possibility of PPh.

4.1 Borrowing processes and the sociocultural history of the Philippines

This ties into the second aspect, which highlights that dismissing the likelihood of diffusion via trade so readily may not be a prudent approach. Reid (2018) argued for this possibility, but Blust (2020: 185) countered that lexical borrowing in trade should favour the types of items which are exchanged in commercial transactions, and that the majority of items in the 1,259 item list are not of this type. We find this point unconvincing, as "trade" is not necessarily restricted to commercial transactions. Long distance trade-type relationships which do not involve narrowly commercial exchange are attested for many linguistic areas, such as the Kula circle in western Melanesia (Malinowski 1922). Long distance interaction of other types is also an important factor in loanword diffusion (Haspelmath & Tadmor 2009).

Archaeological evidence and early Chinese documents have shown that long-distance interaction networks are of great antiquity in the Philippines, that chiefdoms of considerable political complexity were already well developed in 11th Century C.E., and that the initial period of complex society

formation spanned the period from the late first millennium B.C.E. to the early first millennium C.E. (Junker 1998). Spanish accounts from the early colonial period (Loarca 1903; Junker 1998) also give us a glimpse into a world of great political complexity, with smaller and larger polities, speaking mutually unintelligible languages, engaging in constant warfare, alliance building, and tributary relationships. Additionally, maritime trade within the archipelago is well established in the archaeological record for the Philippines, both during and prior to the initiation of "foreign" trade late in the first millennium A.D. (Hutterer 1977; Junker 1990a, 1990b).

Given this complex sociopolitical history, we would argue that the presence of a large amount of loanword diffusion within the Philippine archipelago after the breakup of a higher-level dialect network is more likely than not, and that due to the difficulty alluded to above of distinguishing borrowings from common inheritance in Philippine lexicons, we should be circumspect about accepting arguments for PPh which turn on the quantity of lexical evidence.

Again, although we cannot completely discount the possibility of common inheritance as the source of these etyma, their semantic properties seem to indicate that they could just as easily be loanwords, and in the absence of diagnostic shared sound change we cannot discard either possibility. In our view, this renders the proposed lexical evidence for PPh problematic.

4.2 Geographical and subgroup distribution of PPh etyma

Another piece of evidence that seems to point to the operation of large-scale borrowing processes is the distribution of many of the PPh etyma. In cases such as *lánut, for which Blust's reconstructed PPh meaning is "abaca fibre, fibre-yielding plant, tree with a bark which yields a vine-like fibre", all of the attested reflexes of this form are present only within the Philippine archipelago, with none in the Greater Central Philippine (GCP) languages of northern Sulawesi such as Tondano and Gorontalo, whereas reflexes of this term do in fact appear in the GCP languages Central Subanon, Western Subanon, Aklanon, Mansakan, and Bisayan. In fact, a large number of the etyma which are proposed to be reconstructable to PPh share this pattern of attestation: present in the Philippines proper, but not in the GCP languages of northern Sulawesi. In fact, only seven proposed PPh etyma are found in these languages: *usauR, *láyug, *liqed, *iqit, *habél, *buál, and *butí.

If it was indeed the case that these etyma were present in PPh, we are forced to assume that for some reason all reflexes of these cognate sets were lost or replaced in the GCP languages of northern Sulawesi. A more parsimonious explanation would be that these etyma represent loanwords, which spread via the Philippine maritime interaction sphere, but did not reach the GCP languages in northern Sulawesi further across the ocean. As for the seven PPh etyma which do appear in northern Sulawesi, these could represent either loans which spread throughout the early PMP dialect network, or retentions from PMP for which there are no longer any Philippine-external witnesses. We consider this to be counterevidence for the PPh hypothesis and evidence for an interaction sphere postdating the departure of the GCP North Sulawesi languages for their current location. We will discuss below an example of probable morphosyntactic borrowing which appears to show the same distribution, further reinforcing this interpretation.

The patterns of distribution of other words in the lexemes reconstructed for PPh also indicate that they are more likely to be the product of borrowing, rather than common inheritance. For example, several lexemes show a distributional pattern like that for the proposed PPh protoform *labas 'pass by, overlook something when searching; to pass by (of time)' (Figure 1)

In this type of distributional pattern, reflexes of a proposed PPh lexeme appear in a single large, regional language and several languages spoken by smaller groups. For *labas, the reflex *ag-pa-labás* appears in Ilokano (Northern Luzon microgroup), a major Philippine language of northwestern Luzon

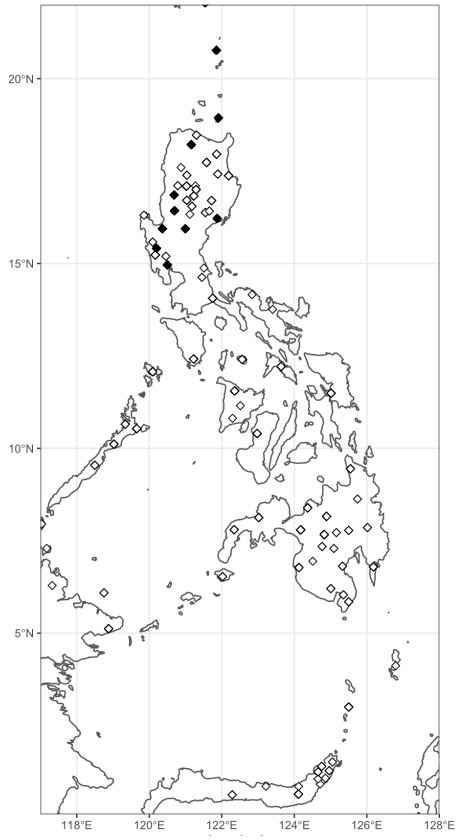


Figure 1: Reflexes of protoform *labas

with around nine million speakers, and which has been an important regional language and lingua franca for many centuries (Rubino 1997). All of the other reflexes of *labas appear in smaller languages also spoken in northern and central Luzon such as Kankanaey, Ibaloi, Casiguran Dumagat (Northern Luzon microgroup) and several negrito languages of central Luzon such as Kampampagan and Ayta Abellen (Central Luzon microgroup). The only other place where reflexes of this term appear is in the Batanes, in the Batanic languages Yami and Itbayat, of which Itbayat is known to have a historical relationship with Ilokano speakers (Maree 2005; Gallego 2020).

The geographical distribution, limited to northern and central Luzon, and the fact that Ilocano has historically been a prestige language in this region strongly suggest that rather than being reconstructable to a deeper protolanguage like PMP or PPh, this term originated in Ilocano and then spread as a loan among smaller languages which it was in contact with. The fact that Ilocano as a regional prestige language and a lingua franca would have been part of the language inventory of bi- or multilingual speakers of smaller groups in this region means that here we have a classic example of a situation with a sociological gradient through which loanwords must have spread from Ilocano into the smaller languages. Even though the meanings of the reflexes of this term do not necessarily fall into a "highly borrowable" category, we know that such situations can lead to lexical replacement in basic vocabulary and closed word classes (Thomason & Kaufman 1998; Haspelmath & Tadmor 2009).

4.3 Further remarks on the lexical evidence

We have shown that rather than pointing to a Proto-Philippines which replaced PMP daughter languages in the Philippines, the lexical evidence in concert with archaeology tells a more interesting story, one of a rich and complex history of contact and cultural change. The compilation of the 1,259 item list of

	Reflex	Reflex Gloss	ACD Name	Speakers, approx.
1	labah	to pass by	Ayta Abellan	3500
2	lábis	excessive, more than enough (as a pole that is longer than it needs to be); excess	Casiguran Dumagat	610
3	debas-en	to make, take , etc. something too far (as house dimension beyond the specifications, bananas beyond proper ripeness)	Ibaloy	116000
4	h<om>abas	to pass by (object, day or time)	Ibatan	33000
5	ag-pa-lábas	to let pass; tolerate; be understanding	Ilokano	8100000
6	na-lábas	past	Isneg	40000
7	pa-xavas-en	to let pass	Itbayaten	3500
8	na-labás	gone; gone away; passed; passed on	Kankanaey	240000
9	labas	pass by, pass through (in the process of leaving, going out)	Kapampangan	2800000
10	on-labás	to go beyond, pass through; surplus, excess above requirements	Pangasinan	1800000
11	ni-mi-avas	passed by	Yami	4000

Table 1: Reflexes of *labas

Philippine-only etyma and their proposed reconstructions is a major resource for uncovering more about Philippine prehistory, and further careful examination should reveal many more interesting details about the Philippine interaction sphere.

5 Insights from morphosyntax: Contact-induced borrowing in CV morphology

We turn now to an underexplored morphosyntactic variation that lends further empirical support to extensive contact across Philippine islands. As will be seen in this section, this variation is widely attested in languages spoken on the major islands of the Philippines, and is absent in outliers that belong to the same linguistic subgroups. This distribution reinforces the view that the reported lexical items found exclusively in Philippine languages are more likely to be an outcome of contact. Most Philippine languages feature a complex voice system inherited from Proto-Austronesian. This voice system is known in the literature as Philippine-type voice, featuring four-way affixal morphology on the verb that controls the argument-marking pattern and extraction restriction of a given clause. See below an example from Tagalog (6a–d). The syntactic pivotal argument in each voice is italicized in the translation. To remain

theory-neutral, the case-markers of the arguments are glossed with abstract labels (CM_1; CM_2; P_1; P_2) as they do not affect the main evidence to be discussed here.

(6) Tagalog

a. B⟨**um**⟩ili si AJ ng keyk mula kay Lia para kay Joy.
buy⟨**AV**⟩ [PN.PIVOT AJ] ID.CM_2 cake P_1 PN.CM_2 Lia P_2 PN.CM_2 Joy
'*AJ* bought cake from Lia for Joy.' (ACTOR VOICE)

b. Bi-bilih-**in** ni AJ ang keyk mula kay Lia para kay Joy.
CONT-buy-**PV** PN.CM_1 AJ [PIVOT cake] P_1 PN.CM_2 Lia P_2 PN.CM_2 Joy
'AJ will buy *cake* from Lia for Joy.' (PATIENT VOICE)

c. Bi-bilih-**an** ni AJ ng keyk si Lia para kay Joy.
CONT-buy-**LV** PN.CM_1 AJ ID.CM2 cake [PN.PIVOT Lia] P_2 PN.CM2 Joy
'AJ will buy cake from *Lia* for Joy.' (LOCATIVE VOICE)

d. **I**-bi-bili ni AJ ng keyk mula kay Lia si Joy.
cv-CONT-buy PN.CM1 AJ ID.CM2 cake P_1 PN.CM2 Lia [PN.PIVOT Joy]
'AJ will buy cake from Lia for *Joy*.' (CIRCUMSTANTIAL VOICE)

A similar four-way voice system is found across higher-order Austronesian languages spoken in Taiwan and the Philippines. Consider below examples of parallel examples of voice alternation in Paiwan (Formosan) and Kimaragang (Dusunic, Malayo-Polynesian).

(7) Paiwan

a. Q**<m>**alup a caucau tua vavuy i gadu tua vuluq.
<AV>hunt [PIVOT man] CM_2 pig LOC mountain CM_2 spear
'The man hunts wild pigs in the mountains with a spear.' (ACTOR VOICE)

b. Qalup-**en** nua caucau a vavuy i gadu tua vuluq.
hunt-**PV** CM_1 man [PIVOT pig] LOC mountain CM_2 spear
'The man hunts wild pigs in the mountains with a spear.' (PATIENT VOICE)

c. Qalup-**an** nua caucau tua vavuy a gadu tua vuluq.
hunt-**LV** CM_1 man CM_2 pig [pivot mountain] CM_2 spear
'The man hunts wild pigs in the mountains with a spear.' (LOCATIVE VOICE)

d. **si**-qalup nua caucau tua vavuy i gadu a vuluq.
CV-hunt CM_1 man CM_2 pig LOC mountain [PIVOT spear]
'The man hunts wild pigs in the mountains with a spear.' (Ferrell 1979:202) (CIRCUMSTANTIAL VOICE)

(8) Kimaragang

a. **Mang**-alapak okuh do niyuw.
AV.TR-split [1SG.PIVOT] CM_2 coconut
'I will split the coconut(s).' (ACTOR VOICE)

b. Lapak-**on** kuh it niyuw.
split-**PV** 1SG.CM_1 [PIVOT coconut]
'I will split the coconut(s).' (PATIENT VOICE)

c. Lapak-**an** kuh do niyuw it wogok.
split-**LV** 1SG.CM_1 CM_2 coconut [PIVOT pig]
'I will split some coconuts for the pigs (to eat).' (BENEFICIARY VOICE)

d. Nokuroh.tu n-**i**-lapak nuh do niyuw inoh dangol kuh?
why PST-**CV**-split 2SG.CM$_1$ CM$_2$ coconut [MED.PIVOT] bush.knife 1SG.POSS
'Why did you use my bush knife to split coconuts?' (Kroeger 2005; glossed ours) (CIRCUMSTANTIAL VOICE)

As shown above, each of the four clauses exhibits a distinct yet unified argument-marking pattern across the three languages (6)–(8). Where the sentence is marked in Circumstantial Voice (CV), the instrument or benefactive bears a special marking labeled as pivot and renders the sole phrase in the clause eligible for relativization ((6d), (7d), and (8d)). There has been a consensus in the literature that this four-way voice morphology can be traced back to Proto-Austronesian (or at least to an early stage of Austronesian prehistory when Austronesian speakers were still in the homeland, Taiwan). As seen in (9), at this early stage, each of the four voices was marked by a single affix.

(9) Early Austronesian voice morphology (Blust 2009; Ross 2009, 2012; Blust & Chen 2017)
a. Actor Voice: *<um>
b. Patient Voice: *-in
c. Locative Voice: *-an
d. Circumstantial Voice: *Si-/Sa-

In the majority of Austronesian primary branches, Circumstantial Voice is free to combine with pivot phrases of a wide range of thematic roles: instrument, theme, benefactive, reason, and cause. Consider below examples from Seediq (10), an Atayalic language spoken in central Taiwan, and Paiwan (11), a single-member primary of Austronesian. Each of the two languages represents a distinct primary branch.

(10) Seediq (Formosan)
a. [s]-hanguc=mu ∅ sari ka dakis.
CV-cook=1SG.CM$_1$ CM$_2$ taro PIVOT Dakis
'I cooked meat for Dakis.' (Chen 2017:101) (benefactive pivot)
b. ga=na [s]-sebuc ∅ ricah ka qreti.
PROG=3SG.CM$_1$ CV-hit CM$_2$ plum PIVOT stick
'He/she is knocking down plums (from the trees) with a stick.' (Chen 2017:99) (instrument pivot)
c. [**s**]-k<n>-narux na temi ka knrudan=na.
CV-STAT<PFV>sick CM$_1$ Temi PIVOT age=3SG.POSS
'Temi got sick because of her age.' (Chen 2017:79) (reason pivot)
d. Wada=mu [**s**]-paadis ∅ dakis ka tigami.
PFV=1SG.CM$_1$ **CV**-send CM$_2$ Dakis PIVOT letter
'I sent Dakis a/the letter.' (Chen 2017:121) (theme pivot in ditransitives)
e. [s]-p-seeliq=mu ∅ robo ka rodux nii.
CV-CAUS-butcher=1SG.CM$_1$ CM$_2$ Robo PIVOT chicken this
'I asked Robo to butcher this chicken.' (theme pivot in causatives)

(11) Paiwan (Formosan)
a. 'u-[s<in>i]-pangul sa a'-pungdjuq ta kasiw ti kapi.
1SG.CM$_1$-CV<PFV>hit LK STAT-broken CM$_2$ wood PIVOT Kpi
'I hit the wood broken for Kepi.' (Wu 2013:192) (benefactive pivot)
b. [si]-tekel ni kapi ta vava a kupu.
CV-drink PN.CM$_1$ Kapi CM$_2$ wine PIVOT cup
'Kapi drinks wines with the cup.' (Wu 2013:31) (instrument pivot)

c. s<in>i-kan m zepul ta ci'aw a za vengaLay nimadu.
CV<PFV>eat PN.CM$_1$ Zepul CM$_2$ fish PIVOT that pregmancy 3SG.CM$_1$
'Zepul ate fish because of her preagnancy.' (Chang 2006:73) (reason pivot)

d. si-pavai ti cemedas a zua hana tjay zapul.
CV-give PN.CM$_1$ Cemedas PIVOT that flower CM$_2$ Zepul
'Cemedas gave that flower to Zepul.' (Chang 2006:297) (theme pivot in ditransitives)

e. 'u-si-pa-veli=anga tjay kapi a vatu.
1SG.CM$_1$-CV-CAUS-sell=COS CM$_2$ Kapi PIVOT dog
'I have caused Kapi to buy (i.e. sold Kapi) the dog.' (Wu 2013:33) (theme pivot in causatives)

Unlike Formosan languages such as Seediq and Paiwan, many Philippine languages exhibit an innovative affix that co-occurs with CV morphology, where the affix specifies the thematic role of the pivot phrase. In Tagalog, for example, the sequence *i-ka-* indicates reason phrases, *i-pag-* indicates instrument, and i-pang- for benefactive. Regardless, there are also verbs that do not take the innovative affix *ka-*, *pag-* and *pang-*.

(12) Tagalog (Malayo-Polynesian)

a. i-p<in>ag-luto ni Kyla ng adobo si Juan.
CV-PAG<PFV>-cook CM$_1$ Kyla INDF.CM$_2$ adobo PIVOT Juan
'Kyla cooked adobo for Juan.' (CV + *pag-* for benefactive pivot)

b. i-p<in>ang-ka-kain ni kyla ang kutsara.
CV-PANG<PFV>-RED-eat PN.CM$_1$ Kyla PIVOT spoon
'Kyla is eating with the spoon.' (CV + *pang-* for instrument pivot)

c. Ang paninigarilyo ang i-k<in>a-matay ni Juan.
PIVOT cancer PIVOT CV-KA<PFV>-die CM$_1$ Juan
'The reason Juan died was because of cancer.' (CV + *ka-* for reason pivot)

On the other hand, other types of pivot possible in CV, such as the theme, do not appear with the additional affix *pang-*, *pag-*, or *ka-*. Consider (13).

(13) Tagalog (Malayo-Polynesian)

a. i-b<in>igay ni juan ang pera kay aya.
CV-<PFV>give PN.CM$_1$ Juan PIVOT money PN.CM$_2$ Aya
'Juan gave Aya the money.' (CV for theme pivot in ditransitives)

b. i-p<in>a-kanta=ko kay ivan ang kanta.
CV-CAUS<PFV>sing=PN.CM$_1$ PN.CM$_2$ Ivan PIVOT song
'I asked Ivan to sing a song.' (CV for theme pivot in causatives)

Itbayaten, a Bashiic language that does not bear a particularly close relationship with Tagalog, exhibits the same four-way thematic-oriented subdivision in CV morphology. Consider (14).

(14) Itbayaten (Malayo-Polynesian)

a. i-cha-hakey
CV-KA-like
'A is the cause of liking.' (Yamada 2015:50) (CV + *ka-* for reason pivot)

b. i-pang-among
CV-PAN-fish
'to fish with A' (Yamada 2015:50) (CV + *paN-* for instrument pivot)

c. i-pang-xap
CV-PAN-get
'to get for A' (Yamada 2015:50) (CV + *paN-* for benefactive pivot)

Notably, similar to Tagalog, not all CV-marked verbs in Itbayaten employ the innovative affix. For example, although some instrument-selecting CV verbs take the affix *pang-* (14b), there are similar verbs that do not require that affix. Consider two reported examples without the innovative affix (15a–b).

(15) Itbayaten (Bashiic, Malayo-Polynesian)

a. i-chali
CV-dig
'to dig with A' (Yamada 2015:50) (CV for instrument pivot)

b. i-'inom
CV-drink
'to drink with A' (Yamada 2015:50) (CV for instrument pivot)

Our survey reveals a similar use of the additional, innovative morphology in languages across different regions of the Philippines. Consider below examples from Asi (Greater Central Philippines) and Pangasinan (Meso-Cordilleran).

(16) Asi (Greater Central Philippines, Malayo-Polynesian)

a. i-pang-limpyo nako kag suka it mga gaha.
F-CV-PANG-clean 1SG.CM$_1$ CM$_2$ vinegar G PL window
'I'll clean the windows with vinegar.' (Hendrickson & Kilgour 1985:39) (CV + *pang-* for instrument pivot)

b. i-sandrok nako sida it suya.
CV-dish.food 1SG.CM$_1$ 3SG.PIVOT G viand
'I'll dish up some of this viand for her.' (CV for benefactive pivot)
(Hendrickson & Kilgour 1985:40)

c. Kag i-k<in>a-matay nida ay kanser.
N CV-KA<PFV>-die 3SG.CM$_1$ PIVOT cancer
'What he died of was cancer.' (Hendrickson & Kilgour 1985:40) (CV + *ka-* for reason pivot)

(17) Pangasinan (Meso-Cordilleran, Malayo-Polynesian)

a. i-mpan-katli nen Mark iyan katli ed samay papel.
CV-PAN.PFV-cut CM$_1$ Mark DEM.PIVOT scissors CM$_2$ cm$_2$ paper
'Mark used this scissors to cut that paper.' (Rosario 2017:81) (CV + PAN- for instrument pivot)

It is important to note that this innovation is not attested in all Philippine languages. While the languages discussed above belong to four of the nine first-order branches under the putative Philippine subgroup, many other members of these branches do not share this innovation. As exemplified in the Karao data (18a–b), these languages either employ a single (retentive) affix *i-* for CV clauses, as do Formosan languages (10)-(11), or employ some other affixes to denote specific types of CV constructions (e.g. *i-...-an* for clauses with a benefactive pivot, as in (18b)).

(18) Karao (Meso-Cordilleran, Malayo-Polynesian)

a. ’i-tegteg na nga’ngi-’i batho-cha ’aramdi.
CV-flatten CM_1 child-PIVOT rock-CM_2 wire
‘The child will use the rock to flatten the wire.’ (Brainard 1997:101) (CV for instrument pivot)

b. ’i-tongkal-an na to’o-’i nga’nga na ’amayo.
CV-buy-AN CM_1 person-PIVOT child CM_2 toy
‘The person will buy the child the toy.’ (Brainard 1997:100)(CV + *-an* for benefactive pivot)

A closer look at the distribution of languages with those innovative affixes reveals two important facts. First, while this innovation is found across the three major islands, Luzon, Visayas, and Mindanao, it is attested neither in Palawan nor in members of the Greater Central Philippine group that are spoken in northern Sulawesi and islands between Sulawesi and Mindanao. Consider the examples below from Talaud (19a–c).

(19) Talaud (GCP, Malayo-Polynesian)

a. lama’a i-saraing ngimangitou.
dish CV-dance 3PL.CM_1
‘Dishes will be used by them in a dance.’ (Utsumi ms.:20) (CV for instrument pivot)

b. inassa n-i-laha=ku huna m-maria.
fish PST-CV-cook=1SG.CM_1 for CM_1-Maria
‘I cooked fish for Maria.’ (Utsumi ms.:24) (CV for benefactive)

c. ana?itou n-i-luass i-piteres [uauggu na-?angkat=te huru].
child 3SG.CM_1 PST-CV-be.pleased [CM_1-Peter AV.PST-promote=COMP teacher]
‘Her/his child made her/him pleased because (her/his child) became a teacher.’ (Utsumi ms:25) (CV for reason pivot)

This distribution indicates that the innovation in CV morphology is unlikely to be an outcome of inheritance from the tree top (i.e. retention from the putative shared common ancestor), as that scenario will entail the innovation being lost in (i) all languages spoken outside the three major islands, (ii) all Negrito languages, and (iii) the majority of languages under the branches attested with this change. Instead, it reinforces the presence of horizontal transfer across the main islands of the Philippines. This innovative CV morphology fits well with Type III structural change (20) defined in Heine & Kuteva (2005:124).

(20) *Type III Structural effect of contact-induced grammaticalization*

The new and the old categories coexist side by side, but the structure of the old category is redefined as a result of the presence of the new category (differentiation).

Given the consensus in the literature that structural borrowing presumes extensive lexical borrowing from the same language (e.g., Thomason 2001; Heine & Kuteva 2005; Aikhenvald 2006; Matras 2009; a.o.), the presence of structural borrowing of the innovative CV morphology is a strong indicator that extensive lexical borrowing would have also taken place across the main Philippines islands. This understudied locus of variation morphosyntax thus lends further empirical support for the lexical diffusion account for the shared lexical items.

6 Conclusion

In this paper, we have re-examined recent arguments for Proto-Philippines as the sole surviving PMP descendent on Luzon (Blust 2019, 2020, 2022; Zorc 1986, 2020) and presented new evidence for an alternative view: there is little motivation for postulating this alleged ancestor, as the claimed evidence for PPh may be explained as the outcome of various layers of diffusion that also involves extensive contact, as suggested previously in Ross (2020). Three lines of new evidence supported to this view. First, the absence of PMP *d/z merger in the Central Luzon subgroup, which undermines the merger as an innovation defining PPh. Second, the geographical distribution of these innovations, along with the semantic categories of reported lexical innovations under Haspelmath & Tadmor's (2009) criteria and both indicate a high likelihood of borrowing (and not inheritance). Third, the distribution of an underexplored variation of Circumstantial Voice (CV) morphology suggests multiple layers of borrowings across Philippine subgroups. This lends new empirical evidence for massive contact on the main islands of the Philippines. We conclude accordingly that there is no obvious motivation to assume the existence of PPh.

7 References

Blake, F. (1906). Contributions to Philippine grammar. *Journal of the American Oriental Society* 27, 317—396.

Blust, R. (1991). The Greater Central Philippine hypothesis. *Oceanic Linguistics* 30(2), 73–129.

Blust, R. (2005). The linguistic macrohistory of the Philippines: Some speculations. In H. Liao & C. Rubino (eds.), *Current topics in Philippine linguistics and anthropology, Parangal kay Lawrence A. Reid*: 31–68. Linguistic Society of the Philippines.

Blust, R. (2019). The resurrection of Proto-Philippines. *Oceanic Linguistics* 58(2), 153–256.

Blust, R. (2020). Response to comments on "The resurrection of Proto-Philippines." *Oceanic Linguistics* 59(12), 450–479.

Charles, M. (1974). Problems in the reconstruction of Proto-Philippine phonology and the subgrouping of the Philippine languages. *Oceanic Linguistics 13*, 457–509.

Ebarhard, D. M., Simons, G. F., & Fennig, C. D. (2021). Ethnologue: Languages of the world.

François, A. (2014). Trees, waves and linkages: Models of language diversification. In C. Bowern & B. Evans (eds.), *The Routledge Handbook of Historical Linguistics*: 161–189. Routledge.

Gallego, M. K. S. (2020). Ibatan of Babuyan Claro (Philippines) - Language contexts. *Language Documentation and Description*, 17, 87-110.

Gray, R. D., Drummond, A. J., & Greenhill, S. J. (2009). Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323(5913), 479–482. https://doi.org/10.1126/science.1166858
Greenhill, S. J., & Gray, R. D. (2009). Austronesian language phylogenies: Myths and misconceptions about Bayesian computational methods. In L. Sagart et al. (eds.), *Austronesian historical linguistics and culture history: A festschrift for Robert Blust*: 1–23). Australian National University.

Haspelmath, M., & Tadmor, U. (2009). The Loanword Typology project and the World Loanword Database. In M. Haspelmath & U. Tadmor (eds.), *Loanwords in the world's language: A comparative handbook*: 1–34). De Gruyter.

Liao, H. (2011). On the development of comitative verbs in Philippine languages. *Language and Linguistics* 12, 205–237.

Liao, H. (2020). A reply to Blust (2019) The resurrection of Proto-Philippines. *Oceanic Linguistics* 59(12), 426–449.

Llamzon, T. A. (1975). Proto-Philippine phonology. *Archipel* 9, 29–42.

Malinowski, B. (1922). *Argonauts of the Western Pacific: An Account of Native Enterprise and Adventure in the Archipelagos of Melanesian New Guinea.* London: Routledge & Kegan Paul.

Maree, J. (2005). *The Ibatan: A genealogy of the people of Babuyan Claro Island.* Summer Institute of Linguistics

Pawley, A. (2006). Explaining the aberrant Austronesian languages of southeast Melanesia: 150 years of debate. *Journal of the Polynesian Society* 115, 215–258.

Paz, C. (1981). *A reconstruction of Proto-Philippine phonemes and morphemes*. Cecilio Lopez Archives of Philippine Languages and the Philippine Linguistics Circle.

Reid, L. (1978). Problems in the reconstruction of Proto-Philippine construction markers. In S. Wurm & L. Carrington (eds.), *Second International Conference on Austronesian Linguistics: Proceedings* Vol. 1: 33–36). Department of Linguistics, Research School of Pacific Studies, Australian National University.

Reid, L. (1982). The demise of Proto-Philippines. In A. Halim, L. Carrington, & S. Wurm (eds.), *Papers from the 3rd International Conference on Austronesian Linguistics* vol. 2: *Tracking the travelers*: 201–216. Australian National University.

Reid, L. (2017). Revisiting the position of Philippine languages in the Austronesian family. The Br. Andrew Gonzalez FSC (BAG) Distinguished Professorial Chair Lecture, 10022.

Reid, L. (2020). Response to Blust "The Resurrection of Proto-Philippines." *Oceanic Linguistics* 59(1–2), 374–393.

Ross, M. (1995). Some current issues in Austronesian linguistics. In *Comparative Austronesian Dictionary*: 45–120. Mouton de Gruyter.

Ross, M. (2005). The Batanic languages in relation to the early history of the Malayo-Polynesian subgroup of Austronesian. *Journal of Austronesian Studies* 1(2), 1–24.

Ross, M. (2020). Comment on Blust "The resurrection of Proto-Philippines." *Oceanic Linguistics* 59(1–2), 366–373. https://doi.org/10.1353/ol.2020.0016

Scheerer, O. (1918). Outlines of the history of exploration of the Philippine languages and their relatives in East and West. *The Philippine Review* 3(1), 59–67.

Smith, A. D. (2017). The Western Malayo-Polynesian Problem. *Oceanic Linguistics* 56(2), 435–490.

Zorc, R. D. (1986). The genetic relationships of Philippine languages. In P. Geraghty, L. Carrington, & S. Wurm (eds.), *FOCAL II: Papers from the 4th International Conference on Austronesian Linguistics*: 147–173. Pacific Linguistics.

Zorc, R. D. (2020). Reactions to Blust's "The resurrection of Proto-Philippines." *Oceanic Linguistics* 59(1–2), 394–425.